Random Forest Combined with Boruta Feature Selection for Tropical Cyclone Classification

Truong-Vinh Le[1,2]、Yuei-An Liou[1]

(1)Center for Space and Remote Sensing Research, National Central University、(2)Taiwan International Graduate Program – Earth System Science Program, Academia Sinica

Tropical cyclones (TCs) are one of the most dangerous meteorological phenomena to threaten the human lives, property, and societies. To determine the impact of a tropical cyclone, it is necessary to classify its strength. This study aims to categorize the tropical cyclone intensity in the Northwest Pacific using a supervised machine learning model called Random Forest (RF) based on thirteen influence factors. Additionally, Boruta feature selection was applied to select the essential attributes of TCs before inputting them into the RF model. The accuracy of the machine learning model was validated by the Saffir-Simpson Hurricane Wind Scale measurements. The findings reveal that RF with optimal hyperparameters (i.e., ntree = 1000, mtry = 2) achieved the best performance in the confusion matrix, accuracy, and Kappa coefficients. According to the model, maximum wind speed, central pressure, latitude, and longitude are the key factors in tropical cyclone classification. Finally, we used the model to illustrate the spatial distribution of historical tropical cyclone classification. The machine learning model coupled with the Boruta feature selection algorithm is an effective strategy for improving the intensity predictive models.

Keywords:Boruta feature selection, machine learning, Northwest Pacific, tropical cyclones